
O TODO É MAIOR DO QUE A SOMA DAS PARTES: ESTADO DA ARTE SOBRE INDEXAÇÃO POR ASSUNTOS NA CRIAÇÃO DE CATÁLOGOS COLETIVOS

The whole is greater than the sum of its parts: a state-of-the-art about subject indexing at the creation of integrated union library catalogs

Jorge Revez

Universidade de Lisboa, Faculdade de Letras, Alameda da Universidade, 1600-214 Lisboa, jrevez@campus.ul.pt

Resumo

Este trabalho investiga o problema da indexação por assuntos no momento da criação de catálogos bibliográficos coletivos centralizados, isto é, o que sucede, numa fusão eletrónica, com os termos inseridos nos campos de descrição dos conteúdos. Procura-se sistematizar neste estado da arte as soluções encontradas, no contexto dos sistemas de organização da informação, para rentabilizar o investimento na utilização de descritores de assunto em catálogos bibliográficos locais no momento da formação/fusão de catálogos coletivos. Conclui-se que, não existindo um modelo que se possa aplicar uniformemente a todos os projetos, as diferentes soluções devem resultar de um esforço de planeamento coletivo que implemente um conjunto adequado de medidas. É proposta uma recomendação para as redes que irão promover a criação de catálogos coletivos centralizados em Portugal.

Palavras-Chave: Indexação por assuntos; Catálogos coletivos; Controlo do vocabulário; Catálogos bibliográficos; OPAC; Sistemas de organização da informação

Abstract

This paper investigates the problem of subject indexing at the creation of integrated union library catalogs, more precisely, what happens in an electronic merging process with the terms entered in the content description fields. It seeks to acknowledge the state-of-the-art concerning the solutions found in the context of the information organization systems to value the investment in the use of subject terms in local catalogs at the creation/merger of union catalogs. We conclude that, in the absence of a model that can be applied uniformly to all projects, different solutions should be the result of a collective planning effort to implement an adequate set of measures. It is proposed a recommendation to the networks that will promote the creation of integrated union catalogs in Portugal.

Keywords: Subject indexing; Union catalogs; Vocabulary control; Library catalogs; OPAC; Information organization systems

1 Introdução

A 1 de outubro de 2015, o OCLC imprimiu as últimas fichas catalográficas em papel (OCLC, 2015). Foi o fim de um ciclo na história dos catálogos bibliográficos, anunciado pelas potencialidades e pelos desafios das novas tecnologias de computação em rede: o desenvolvimento dos metadados, a presença dos catálogos na Web e o uso dos protocolos de troca de dados (Olson, 2001, p. 319).

O catálogo bibliográfico é o conjunto dos registos bibliográficos, isto é, da descrição física e intelectual dos documentos existentes numa coleção, fornecendo, entre outros elementos, a sua localização e forma de acesso. Até à massificação do uso da Web, o catálogo era quase a própria biblioteca. Era a única representação das coleções e, por isso, a única forma de

pesquisar a informação. Com a disponibilização dos *Online Public Access Catalogs* (OPAC) todos os catálogos poderiam ser coletivos. Inversamente, todos os catálogos coletivos poderiam ser locais através da filtragem dos registos da própria biblioteca no interface dos utilizadores. Desta forma, os OPAC foram uma mudança de foco, das coleções locais à possibilidade de convergência de múltiplas coleções, acedidas remotamente e sem barreiras geográficas (Borgman, 1997).

Numa síntese recente, Lazarini (2015) assinalou três momentos fulcrais na história dos catálogos bibliográficos e seus objetivos: as Regras de Cutter (1876), os Princípios de Paris (1961) e o modelo conceptual FRBR (*Functional Requirements for Bibliographic Records*) seguido da Declaração da IFLA de 2009 com os novos princípios internacionais

de catalogação. Este percurso é marcado pelo desenvolvimento de inúmeras formas de codificar e normalizar a informação registada nos catálogos, sendo que os elementos considerados como mais suscetíveis de escapar ao controlo bibliográfico são naturalmente os relativos à descrição dos assuntos, a componente intelectual do registo.

No momento da criação de catálogos bibliográficos coletivos - sistemas de organização da informação que suportam os dados bibliográficos de uma rede de bibliotecas - desenvolvem-se operações informáticas complexas. A tarefa dificulta-se mais quando se decide fundir registos bibliográficos que provêm de sistemas antes isolados.

O problema que aqui abordaremos tem uma natureza técnica e política, combinando gestão e tecnologia da informação. Perguntamos então o que fazer com os descritores utilizados pelos catálogos bibliográficos locais no momento da fusão dos registos bibliográficos? Como garantir que o investimento das diferentes bibliotecas na descrição do conteúdo temático é devidamente salvaguardado? Como assegurar que a recuperação da informação por assuntos se mantém intacta após a fusão?

O objetivo geral consiste em apresentar o estado da arte sobre as diferentes teorias e experiências na fusão de registos bibliográficos (em particular, dos campos de assunto) aquando da criação de catálogos coletivos centralizados, ou seja quando é constituída uma base de dados comum para descrever diferentes coleções locais. Pretendemos especificamente reconhecer os principais debates e correntes; identificar experiências e as soluções encontradas no contexto dos OPAC; e apresentar uma síntese da literatura analisada em forma de recomendação para os catálogos a desenvolver em Portugal.

O princípio aristotélico de que “o todo é maior do que a simples soma das partes” é o mote para a abordagem de um problema antigo: procurar que um catálogo coletivo, como um todo, respeite o valor das partes, em toda a sua potencialidade analítica e descritiva. Seguimos uma metodologia qualitativa de natureza exploratória utilizando técnicas de pesquisa e recuperação, com análise da informação bibliográfica, subordinada a dois temas principais: 1) os catálogos coletivos como experiência de organização da informação; 2) a utilização da indexação por assuntos nos catálogos bibliográficos, particularmente no caso dos catálogos coletivos centralizados, beneficiando do ambiente digital.

2 O conceito de catálogo coletivo

Para tentarmos definir catálogo coletivo, seguimos a definição de Reitz (2013) para “union catalog”:

A list of the holdings of all the libraries in a library system, or of all or a portion of the collections of a group of independent libraries, indicating by name and/or location symbol which libraries own at least one copy of each item.

Esta proposta realça a disponibilização da propriedade e localização dos documentos e a fusão desta informação num único interface. Reitz distingue o catálogo coletivo centralizado, que reúne os registos bibliográficos numa única base de dados, da sua versão virtual, que seria apenas um mecanismo de pesquisa federada num único interface, para a recuperação de registos de diferentes bibliotecas, isto é, de diferentes catálogos, através de protocolos como a norma Z39.50 ou outras. Outros autores referem ainda a possibilidade de uma abordagem mista, como por exemplo, o armazenamento dos registos bibliográficos numa base central guardando os dados da circulação (empréstimos, leitores, existências, etc.) nos sistemas locais (Boberić-Krstićev & Tešendić, 2015).

Na verdade, o catálogo coletivo resulta de uma opção política de implementação de trabalho colaborativo (Lass & Quandt, 2004, p. XXIII). Neste sentido, Smith define que os catálogos coletivos devem ser considerados no contexto da catalogação partilhada, das redes e do controlo bibliográfico (2003, p. 643). A interligação das redes com partilha de tarefas e normalização é o produto de um trabalho de união de esforços cooperativos. A razão destes esforços é evidente e tem uma forte motivação política: a coordenação e a gestão de recursos que se querem partilhados, evitando sobretudo a sua duplicação. Como é óbvio, a utilidade de um catálogo está limitada se incluir apenas as existências de uma única biblioteca. A mais-valia decisiva para os leitores é a possibilidade de pesquisar diversas coleções numa dada rede ou região (Smith, 2003, p. 643). A mudança do suporte e da tecnologia utilizada nos registos bibliográficos, com a introdução dos “machine-readable records”, sobretudo a partir da década de 60 do séc. XX, potenciou o desenvolvimento dos catálogos coletivos, a partilha do acesso aos registos e a capacidade dos empréstimos interbibliotecas (Wiley, 2011, p. 200).

Entendemos que o catálogo coletivo é o sistema que permite a recuperação de registos bibliográficos criados em sistemas locais e que pode desenvolver, através de catalogação colaborativa, a descrição das coleções de uma determinada rede de bibliotecas. Nas múltiplas versões possíveis (centralizado, virtual ou misto), a formação de catálogos coletivos deve obedecer a um rigoroso protocolo de instalação e funcionamento. Deve ser claro o seu propósito, o seu âmbito de atuação e os diversos atores envolvidos. A estratégia a seguir pode depender também da dimensão das coleções, das capacidades de pesquisa dos vários tipos de

utilizadores ou do apoio informático disponível para a sua implementação (Dovey, 2000).

No entanto, um dos principais problemas foi sempre o da harmonização, normalização e combinação das linguagens de descrição de assuntos, daí a análise centrar-se nesse momento de fusão, em ambiente Web, e o que ocorre então com os diferentes descritores.

3 A indexação por assuntos nos catálogos coletivos

Existem duas vias-chave e problematizantes para compreendermos o nosso problema: uma delas é a automatização dos catálogos bibliográficos, isto é, a passagem do modelo analógico de catálogo para um modelo eletrónico e, mais tarde, um modelo em rede. A outra é a questão do controlo dos vocabulários que tem uma história que remonta às primeiras formas de codificação simbólica da humanidade.

Começando pela primeira via, Kilgour resenhou esta história desde 1954, destacando a normalização promovida pelos chamados *registos legíveis por máquinas*. Aponta, entre outros aspetos, o trabalho da Biblioteca do Congresso na criação e padronização do formato MARC. Da mesma forma que a padronização do tamanho das fichas bibliográficas em papel havia permitido a troca de registos entre diferentes catálogos, o MARC permitiu a partilha de registos eletrónicos (1970, p. 225).

Este formato representou então o início da possibilidade de formação de catálogos coletivos automatizados; a desconstrução do registo permitiria mesmo à computação ir além das capacidades até então conhecidas: “It seems likely that the immediate future will witness increasing degrees of computerization based on individual words in bibliographic descriptions rather than on the record as a whole” (Kilgour, 1970, p. 226).

Esta transformação já havia sido assinalada por Kuncaitis, que indicou alguns autores que discutiam a possibilidade de se estabelecer um catálogo computadorizado, disponível para ser utilizado por qualquer biblioteca (1968, p. 13). O autor sublinhou a necessidade de se compreender a génese destes sistemas como uma resposta às necessidades de satisfação dos pedidos de empréstimo interbibliotecas. A sua revisão aponta como momento fundador dos catálogos coletivos o ano de 1410, pelo trabalho do monge John Boston de Bury na criação do *Catalog Scriptorum Ecclesiae* (Kuncaitis, 1968, p. 6-7).

Sabemos que o sonho de uma bibliografia universal foi depois perseguido por diferentes autores, com destaque para Paul Otlet e as ideias do Repertório Universal e do *Mundaneum* (Wright, 2014). No entanto, é no início do séc. XX que se assiste ao desenvolvimento de um catálogo coletivo dos Estados Unidos pela Biblioteca

do Congresso. A partir da década de 30, diferentes projetos regionais ganham forma, sobretudo nos Estados Unidos e Canadá, formando uma multiplicidade de redes, em paralelo com o desenvolvimento dos primeiros catálogos coletivos nacionais na Europa Ocidental (Brummel, 1958).

É importante assinalar que a revisão de Kuncaitis, claramente situada num tempo de transição tecnológica, apontava já para a importância do estabelecimento de regras catalográficas que promovessem nos dados bibliográficos recebidos centralmente a uniformização e coerência do produto final (1968, p. 14). A prática catalográfica, incluindo a indexação por assuntos, deveria estar sustentada na experiência de uma biblioteca de grande dimensão. As outras bibliotecas da rede deveriam adaptar as suas práticas a esse centro bibliográfico (Kuncaitis, 1968, p. 12). A escolha da estratégia adotada pela biblioteca-referência da rede, bem como da linguagem documental mais adequada a adotar posteriormente pela rede de bibliotecas, seriam assim duas opções fundamentais.

Kilgour, Long, Landgraf, e Wyckoff (1972) relataram o início do OCLC (Ohio College Library Center, atualmente Online Computer Library Center) em 1967. Este projeto colaborativo, que esteve na origem da maior experiência mundial de catalogação partilhada, e consequentemente do maior catálogo coletivo comercial – *WorldCat* –, foi essencial para a compreensão das virtudes do trabalho cooperativo, bem como da dificuldade de gerir linguagens documentais diversas.

Keller, na sua revisão da situação dos catálogos coletivos nos Estados Unidos e, em particular, na Califórnia, apontou o problema do modelo a seguir para a construção destes sistemas, debatendo a necessidade de uma coordenação nacional dos esforços das redes locais em contraponto com um controlo bibliográfico excessivamente centralizado (1974, p. 2). A discussão sobre o modelo gestor dos catálogos coletivos, ocorrida nos anos 70, teria depois o seu percurso intelectual nas décadas seguintes, nomeadamente em Portugal, não tendo ficado resolvida de forma a oferecer um modelo de sucesso.

Bates faz um balanço destes problemas chamando a atenção para alguns aspetos essenciais: em primeiro lugar, o fato da transição dos catálogos em papel para o formato eletrónico não ter representado uma mudança conceptual: “Online catalogs to date have added powerful capabilities to the traditional catalog, yet system designs, generally, have still not gone beyond implementing the card catalog in online form, with some established online search features tacked on” (1986, p. 357). Para existir essa mudança seria necessário uma mudança do *design* dos sistemas de indexação e de apoio ao utilizador no momento da

pesquisa, que extraísse vantagens das potencialidades informáticas. Este novo modelo facilitaria a relação utilizador-catálogo, mediante a ajuda de um “end-user thesaurus, a front-end system mind linked to existing Library of Congress subject heading indexing, powerful system search capabilities, and additional information in the document records to help in selection” (p. 373).

Destacamos que, neste período, a continuação da utilização dos cabeçalhos de assunto da Biblioteca do Congresso (ou da sua versão abreviada *Sears*) no novo paradigma de catálogos era inquestionável. A existência deste sistema possibilitou que a formação de catálogos coletivos, no que respeita ao problema da indexação, se tornasse mais fácil, em contraste com realidades onde não existiam sistemas com estas características.

Na realidade, estudos acerca do comportamento dos utilizadores concluíram que a pesquisa por assunto nos catálogos informatizados era relevante estatisticamente, ainda que inúmeros problemas surgissem na disponibilização dessa funcionalidade (Connell, 1991, p. 87). Com o desenvolvimento da pesquisa por palavra-chave, muitos autores questionaram a utilidade dos vocabulários controlados no ambiente *Web*. Contudo, Gross e Taylor (2005) mostraram que mais de um terço dos registos recuperados com aquele método de pesquisa se perderia se não existissem cabeçalhos de assunto e, nalguns casos, essa perda poderia atingir os 80-100%. Dez anos depois, replicaram o estudo num catálogo enriquecido com metadados, como sumários e tabelas de conteúdos, e a percentagem de perdas permaneceu alta, o que significava que era necessário preservar o uso dos vocabulários controlados (Gross, Taylor & Joudrey, 2015).

A sua relevância não resolvia, porém o problema do controlo, como afirmaram Lancaster, Connell, Bishop, e McCowan de forma contundente:

Despite popular belief, the transformation of the card catalog into an online database has not significantly improved subject access. Indeed, it may have made the situation worse because it has led to the creation of much larger catalogs that represent the holdings of many libraries. Merging several catalogs into one, when each component catalog provides inadequate subject access, exacerbates the problem, since the larger the catalog the more discriminating must be the subject access points provided (1991, p. 389).

Efetivamente, a fusão dos catálogos pode levantar mais problemas, sobretudo se as bibliotecas não reconhecerem que um catálogo coletivo é algo novo e mais do que a soma das partes. Kohl (1997) reconhecia esse problema e definia-o com precisão:

The first thing the members of a library consortium do is share catalog information. By itself, however, shared in-

formation about collections does not represent the watershed step between cooperation and integration. An integrated catalog is not just the sum of all the catalogs in the consortium but a new creation. An integrated catalog, for example, would not contain multiple bibliographic records for the same item, rather only the best or most complete bibliographic record for an individual bibliographic item is used, with holdings instead of bibliographic records indicating the item's location in various consortial libraries (p. 437).

Reforçando a ideia que já tínhamos avançado anteriormente, Kohl volta a sublinhar a importância dos registos de referência, que servirão de modelos únicos para a eliminação de duplicados. Apesar dos constrangimentos que as variações na forma de catalogar possam trazer a um catálogo coletivo centralizado, Lynch argumenta que é preferível esta solução a um modelo distribuído ou federado (1997, p. 452). Eis outro debate que permanece em aberto.

O problema da coerência nos catálogos coletivos foi igualmente pensado em Portugal, sendo referido por Mendes e Simões:

Dado o volume e o amplo leque temático deste catálogo [Biblioteca Geral da Universidade de Coimbra] - lembramos que teve início em 1958 - a lista final controlada e estruturada, constituirá uma lista de autoridade, a qual poderá servir núcleos documentais de carácter enciclopédico ou a cooperação entre várias bibliotecas de diversas especialidades, como é o caso do Sistema Integrado de Informação Bibliográfica da Universidade de Coimbra (SIIB/UC). Pensamos que este estudo poderá conduzir, através de reflexões conjuntas a uma unidade de pensamento e a uma formação técnica uniforme, permitindo, ainda, a constituição, utilização e permanente actualização de uma linguagem de indexação comum de qualidade, reconhecidamente em falta neste novo catálogo colectivo (2002, p. 14).

Estava em causa a criação de uma lista que servisse como instrumento orientador da indexação por assuntos num ambiente informacional com tão grande quantidade de dados. O instrumento mais utilizado pelas bibliotecas ou bibliografias ocidentais eram os *Library of Congress Subject Headings*, em inglês ou traduzidos (Caro Castro, 2005, p. 55).

Nos anos 90, foram desenvolvidas outras listas de autoridade como o *RAMEAU - Répertoire d'autorité matière encyclopedique et alphabetique unifié* ou o *SIPORbase - Sistema de Indexação em Português* (desde 1988). Neste caso, referente a Portugal, o projeto teve um alcance limitado, ainda que não abundem os estudos de avaliação de indexação que o possam comprovar inequivocamente (uma das poucas exceções é a tese de Martins, 2014). No contexto dos trabalhos normativos e cooperativos gerados pelo projeto PORBASE (Base Nacional de Dados Bibliográficos) foi igualmente desenvolvido o projeto CLIP (Compatibilização de Linguagens de Indexação em Português), coordenado desde 1988 pela Biblioteca

Nacional, com a construção de terminologias nalgumas áreas temáticas, sem lograr o efeito global pretendido.

Procurando ultrapassar as barreiras linguísticas, destaca-se o projeto colaborativo *MACS (Multilingual Access to Subjects)* promovido desde 1997 entre a British Library, a Deutsche Nationalbibliothek, a Bibliothèque nationale de France e a Swiss National Library, com o objetivo de fornecer descritores de assunto multilingues com base nas linguagens LCSH (inglês), SWD (*Schlagwortnormdatei* em alemão) e RAMEAU (francês) (Landry, 2004).

Lancaster (2002) foi um dos autores fundamentais para a sistematização do problema do controlo do vocabulário, publicando em 1972 uma obra clássica sobre o tema: *Vocabulary control for information retrieval*. Barité resumiu recentemente a história do controlo de vocabulários realçando que o impacto da internet se verifica em três direções: a) o surgimento de novos tipos de sistemas controlados ou semi-controlados cuja estrutura interna põe em causa o modelo tradicional de criação de vocabulários controlados (ex. taxonomias Web, folksonomias); b) o desenvolvimento de interfaces e ajudas semi-controlados para a pesquisa, funcionando como novas formas de controlo linguístico (ex. preenchimento automático de campos, “será que quis dizer...”; e c) o desenvolvimento de normas e procedimentos para promover a interoperabilidade (2014, p. 112). Este último ponto interessa-nos particularmente dado que na organização do conhecimento, a interoperabilidade pode ser entendida como a troca de dados entre vocabulários controlados e entre estes e a linguagem natural ou com outros elementos presentes num sistema de informação. Barité questiona mesmo se a interoperabilidade não poderá ser um tipo de controlo de segundo nível que “busca armonizar los procesos de control aplicados a vocabularios contruidos bajo lógicas diferentes, con vistas a favorecer el acceso a grandes continentes de información” como é o caso dos catálogos coletivos, entre outros (p. 114). Vejamos então algumas experiências concretas.

4 Experiências

Foram várias as experiências e os debates identificados na literatura analisada. De uma forma ou de outra, os projetos cooperativos procuraram encontrar soluções para a questão da descrição e da pesquisa por assuntos. Indicaremos apenas quatro experiências e a um caminho em aberto.

Em Espanha, o manual de procedimentos para a inclusão de registos no catálogo coletivo da REBIUN (2014), indica que o número de campos de assunto inseridos é valorizado no momento da eliminação de duplicados, como “critério de conservação do melhor registo”. As 95 bibliotecas cooperantes com a *Red de Bibliotecas Universitarias Españolas* enviam os seus

registos seis vezes por ano formando uma base centralizada com 34 milhões de registos. Não encontrámos nenhuma indicação quanto à existência de regras pré-definidas no consórcio sobre o controlo do vocabulário.

Cousins apresentou a experiência do COPAC (Reino Unido) quanto à metodologia de integração dos registos bibliográficos de diferentes catálogos, focando também o problema da descrição de assuntos. A autora refere que depois da deteção dos duplicados existe um processo de consolidação para formar um registo único: “This consolidation process also merges the subject information found in the contributing records, often resulting in improved subject access” (1999, p. 98). Num exemplo prático:

Any Subject Headings in the subsidiary record which are different from those in the base record are incorporated into the consolidated record. In the example, there is no overlap in the Subject Headings provided by the two records, so all the Subject Headings in the Edinburgh record are included in the new COPAC record (1998, p. 236).

Desde 2009, na Flórida, EUA, 11 bibliotecas universitárias trabalham na fusão dos seus catálogos. Mandatadas politicamente, as três maiores bibliotecas lideraram o processo, beneficiando do serviço de um grupo de trabalho central. Heron et al. (2013) referem que se conseguiu um equilíbrio entre a partilha de dados e o investimento local, o que venceu os receios da perda de qualidade e identidade. Ocorreram depois outros ganhos que muito influenciaram a recomendação que a seguir propomos. Quanto aos assuntos, o grupo de trabalho ponderou abandonar os registos de autoridade locais em favor do uso dos ficheiros da Biblioteca do Congresso, o que se veio a concretizar apesar da resistência de algumas bibliotecas. Os cabeçalhos locais podiam ser introduzidos mas não eram objeto de controlo automático de autoridades (State University Libraries of Florida, 2010).

Koltay apresenta as experiências técnicas levadas a cabo na Hungria, de forma a ultrapassar a presença de múltiplas linguagens controladas. Ao invés de pretender criar uma nova linguagem, persuadindo as diferentes bases de dados a abandonar as suas práticas, a base Matriksz tenta trabalhar com o que está disponível:

It aims at storing the various subject schemes, creates parallelisms among the most important schemes, offers their records for download, thus promoting their use, and develops a search tool that counteracts the confusion created for library users by the existence of several schemes (2004, p. 323).

Podemos concluir que não existe uma forma única de realizar estas operações. As variáveis em presença e o contexto dos sistemas determinam abordagens necessariamente locais que tenham em conta o que

existe, o que se quer fazer e como fazer. Isto implica escolher entre várias opções, o que se torna mais complicado com as muitas variáveis tecnológicas, económicas, normativas e, enfim, políticas.

Um caminho que está em aberto é a interoperabilidade semântica. Zeng e Chan (2004, p. 382-385) analisaram diferentes experiências e sumariaram oito métodos para a realização da interoperabilidade entre linguagens de indexação. Boccato e Torquetti (2012) aplicaram um destes métodos a um serviço de *clipping* com resultados interessantes. Zeng e Chan concluem mesmo que é possível estabelecer uma concordância “among the vocabularies involved or links are built directly among subject authority records of various online catalog systems. Interoperability can be achieved through a variety of approaches” (2004, p. 387-388).

Harken et al. (2006) apresentaram uma definição de interoperabilidade semântica aplicada a assuntos:

The ability of two or more systems or components to exchange or harmonize cognate subject vocabularies and/or knowledge organization schemes to be used for effective and efficient resource discovery without significant loss of lexical or connotative meaning and without special effort by the user (p. 3).

Acrescentaram ainda uma extensa lista de atributos para as soluções deste tipo (p. 35-36), o que serve como um referencial para o desenvolvimento de iniciativas similares, entre outros elementos, que fazem deste relatório uma peça essencial para compreender uma das linhas fundamentais de trabalho no futuro.

Mais recentemente, a IFLA publicou uma extensão do modelo conceptual FRBR para os dados relativos a autoridades de assunto, os *Functional Requirements for Subject Authority Data* (FRSAD). Com o desenvolvimento deste modelo prevê-se que este tipo de dados “will be able to become part of linked open data and contribute to the further development of the Semantic Web” (IFLA, 2010, p. 50).

5 Recomendação

O panorama dos sistemas de organização da informação bibliográfica mudou radicalmente nos últimos 25 anos em Portugal. No início dos anos 90, as experiências de automatização dos catálogos bibliográficos eram incipientes e a maior parte das bibliotecas continuava a alimentar sistemas analógicos com fichas bibliográficas em papel.

Estes anos foram marcados, na generalidade, pela transição destes sistemas em papel para sistemas automatizados que essencialmente reproduziam nos registos eletrónicos a mesma estrutura de descrição e codificação usada no papel. A conversão ocorreu em primeiro lugar ao nível do suporte, do papel para o eletrónico, possibilitando o acesso local à informação disponibilizada através de um equipamento

informático. Mais tarde mudou tecnologicamente, beneficiando do desenvolvimento dos sistemas de redes informáticas, particularmente do desenvolvimento da internet, proporcionando a criação de catálogos bibliográficos acessíveis em redes intranet e na *World Wide Web*.

As redes e as trocas de informação que aquelas possibilitam aceleraram o crescimento dos catálogos coletivos apesar de este tipo de projetos nunca ter atingido em Portugal uma escala verdadeiramente nacional. Nos casos em que foi possível implementá-los ocorreram ganhos imediatos para os utilizadores, sobretudo ao nível da capacidade de recuperação de informação no seio das diferentes redes de bibliotecas.

Os catálogos bibliográficos estão hoje numa encruzilhada (na expressão de Lass & Quandt, 2004). Por um lado, a sua estrutura tradicional de organização da informação, de descrição documento a documento, tem sido questionada e proposta a sua substituição por outros modelos (Coyle, 2016). Por outro lado, os registos bibliográficos têm sido entendidos, no ambiente *Web*, como parte dos grandes conjuntos de dados que, em certos contextos, estão a ser armazenados nos silos dos designados sistemas ou serviços de descoberta - *discovery services* – a par dos registos dos repositórios, das bases de dados de artigos eletrónicos, entre outros (Calhoun, 2006; Han, 2012). A aparente perda de autonomia e relevância das funções do catálogo bibliográfico coexiste com uma revalorização do seu papel de sistema de organização da informação local, como contraponto a uma espécie de *big data* que a qualquer momento se pode transformar num *big chaos* para os utilizadores.

A maioria das bibliotecas portuguesas ainda não chegou a esta encruzilhada. A cooperação ativa e a partilha de dados, fundamentais na formação de catálogos coletivos, constituem um dos grandes desafios dos próximos anos. Vejam-se os casos da rede nacional de bibliotecas públicas e da articulação nacional das bibliotecas do ensino superior – após o insucesso da constituição da RUBI (Rede Universitária de Bibliotecas e Informação, 1996-2000) (Lemos & Macedo, 2003) – ou a situação incipiente das bibliotecas da Administração Pública, três enormes sistemas que não estão interligados nem comunicam entre si os dados bibliográficos. Excetuam-se algumas redes universitárias que trabalham coletivamente há mais de dez anos, como é o caso da Universidade de Lisboa e da Universidade de Coimbra, entre outros exemplos.

Face a este quadro, propomos uma recomendação para a elaboração de projetos de implementação de catálogos coletivos em redes de bibliotecas com sistemas automatizados em funcionamento, com especial enfoque na questão da indexação (Apêndice 1).

Estruturámos a recomendação seguindo a metodologia de Abadal Falgueras (2004), sugerindo que a gestão inicial do projeto seja feita por um grupo-líder sob a coordenação de um dos elementos da rede e com a participação de membros de todas as bibliotecas. O grupo-líder deverá realizar algumas *due diligence*, usando uma expressão da gestão de projetos, a saber:

1. Avaliar se é pertinente levar a cabo um projeto com estas características, definindo uma metodologia de avaliação global que beneficie da combinação de diferentes técnicas de recolha de dados;
2. Produzir documentação inequívoca que possa suscitar um mandato político claro e mobilizador;
3. Obter esse mandato político;
4. Garantir a formação no grupo-base de uma cultura de cooperação ativa, partilha de conhecimento e respeito pelas diferentes idiossincrasias dos elementos da rede.

6 Conclusão

Em síntese podemos afirmar que o estado da arte revela a presença de três modelos (Apêndice 2) que relacionam os catálogos coletivos e a indexação por assuntos. Não pretendemos demonstrar uma linha do tempo de caráter determinista, mas apenas representar o que em alguns contextos se verificou: a coexistência ou sobreposição destas três formas de encarar os sistemas de organização da informação que designámos por catálogos coletivos e muitas dúvidas (representadas pelo ponto de interrogação) respeitantes ao problema da indexação por assuntos.

Em suma, a literatura apresenta cinco grandes núcleos de questões. Em primeiro lugar, é destacado que a automatização, os formatos normalizados e as redes permitiram a comunicação de dados bibliográficos que está na génese dos catálogos coletivos em ambiente Web. Contudo, é consensual, num segundo plano, que não existe um modelo que permita aproveitar todo o investimento realizado pelas bibliotecas locais e simultaneamente evitar a multiplicação de linguagens. Numa terceira linha, o modelo centralizado permite a catalogação em rede, cooperativa, exigindo no entanto o estabelecimento de regras quanto às regras e linguagens documentais a adotar na indexação por assuntos. Daí que a literatura destaque um quarto núcleo onde se argumenta que o modelo centralizado exige uma coordenação que faça o planeamento da criação e desenvolvimento do catálogo coletivo e que estabeleça a forma de fusão dos registos bibliográficos duplicados. Por fim, e em último plano, apontando para o futuro, uma solução possível poderá ser o planeamento atempado da fusão, com o contributo tecnológico e teórico-prático da interoperabilidade das linguagens de indexação.

Estas cinco dimensões denotam, por um lado, a importância dos catálogos e das linguagens documentais que os sustentam; e, por outro, revelam o impacto positivo das novas tecnologias e os desafios e as possibilidades de trabalho que se apresentam futuramente. Conclui-se assim que não existindo um modelo ideal ou uma fórmula consagrada para assegurar o investimento local na descrição de assuntos após a criação dos catálogos centralizados, as diferentes soluções possíveis devem resultar de um esforço de planeamento coletivo que implemente um conjunto adequado de medidas. Procurámos mostrar neste estado da arte – que não pode deixar de ser parcial, dada à quantidade de trabalhos publicados – questionar uma pequena parte de um problema que irá permanecer: a organização do conhecimento.

Referências

- Abadal Falgueras, E. (2004). *Gestión de proyectos en información y documentación*. Gijón: Trea, 2004.
- Barité, M. (2014). Control de vocabulário: orígenes, evolución y proyección. // *Ciência da Informação*. 43:1 (2014) 95-119.
- Bates, M. J. (1986). Subject access in online catalogs: a design model. // *Journal of The American Society for Information Science*. 37:6 (1986) 357-376.
- Boberić-Krstićev, D., & Tešendić, D. (2015). Mixed approach in creating a university union catalogue. // *The Electronic Library*. 33:6 (2015) 970-989.
- Boccatto, V. R. C., & Torquetti, M. C. (2012). Interoperabilidade entre linguagens de indexação como recurso de construção de instrumento de representação temática de clippings de coordenadorias de comunicação social em ambientes universitários: uma proposta metodológica. // *Informação & Informação*. 17:3 (2012) 76-101.
- Borgman, C. L. (1997). From acting locally to thinking globally: a brief history of library automation. // *The Library Quarterly*. 67:3 (1997) 215-249.
- Brummel, L. (1958). Union catalogues: problems of international co-operation. // *Symposium on National Libraries in Europe*. Paris: Unesco, 1958.
- Calhoun, K. (2006). The changing nature of the catalog and its integration with other discovery tools: final report prepared for Library of Congress. <http://www.loc.gov/catdir/calhoun-report-final.pdf> (2015-12-20).
- Caro Castro, C. (2005). *El acceso por materias en los catálogos en línea: índices y terminología de los usuarios en el catálogo Cisne* (Universidad Complutense de Madrid). Salamanca: Ediciones Universidad de Salamanca, 2005.
- Connell, T. H. (1991). Techniques to improve subject retrieval in online catalogs: flexible access to elements in the bibliographic record. // *Information Technology and Libraries*. 10:2 (1991) 87-98.
- Cousins, S. (1998). Duplicate detection and record consolidation in large bibliographic databases: the COPAC database experience. // *Journal of Information Science*. 24:4 (1998) 231-240.
- Cousins, S. (1999). Virtual OPACs versus union database: two models of union catalogue provision. // *The Electronic Library*. 17:2 (1999) 97 – 103.
- Coyle, K. (2016). *FRBR, before and after: a look at our bibliographic models*. Chicago: ALA, 2016.
- Dovey, M. (2000). *So You Want to Build a Union Catalogue?* // *Ariadne*. 23, 2000.

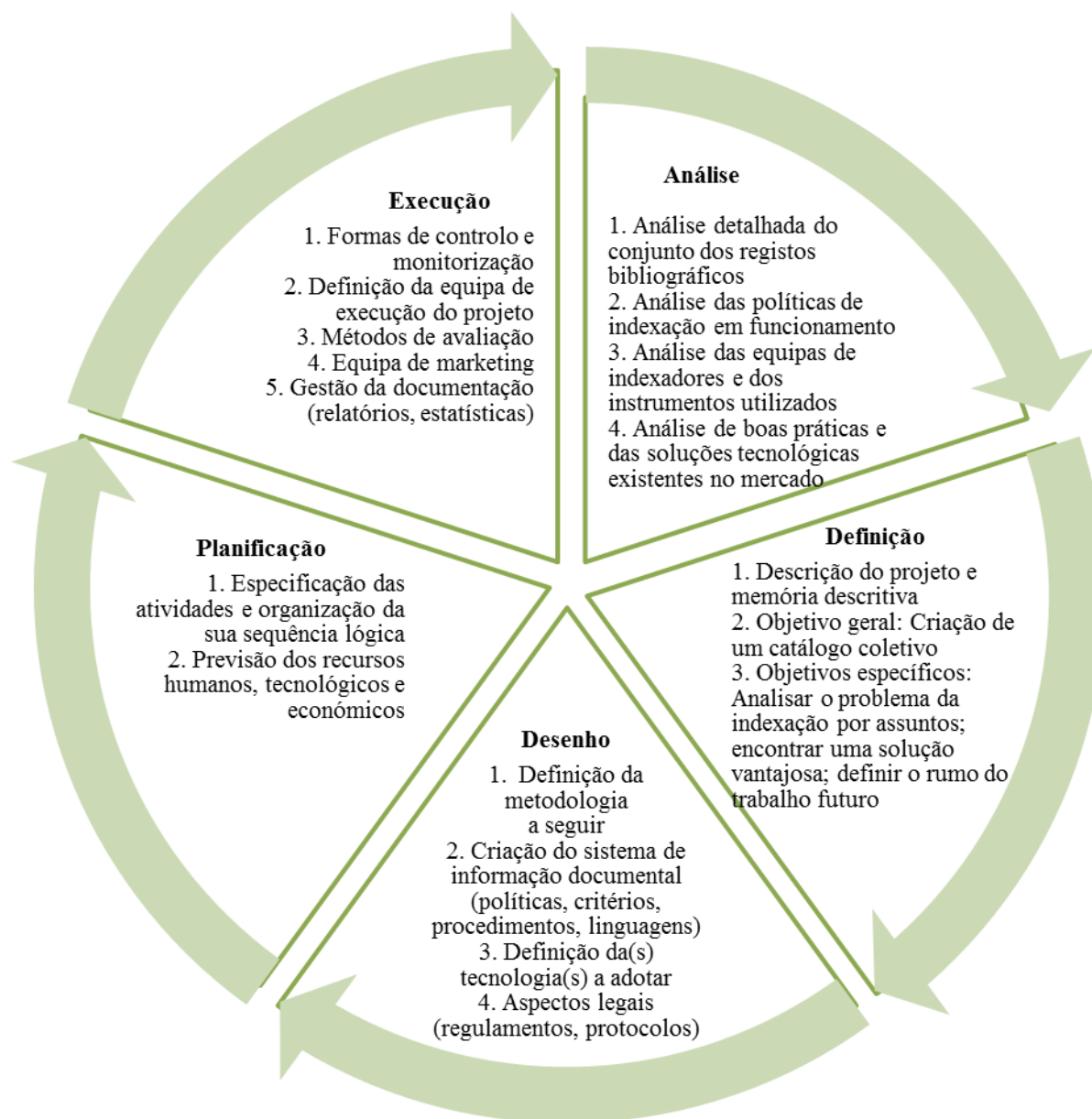
- Gross, T. & Taylor, A. G. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. // *College & Research Libraries*. 66:3 (2005) 212-230.
- Gross, T., Taylor, A. G., & Joudrey, D. N. (2015). Still a lot to lose: the role of controlled vocabulary in keyword searching. // *Cataloging & Classification Quarterly*. 53:1 (2015) 1-39.
- Han, M. J. (2012). New discovery services and library bibliographic control. // *Library Trends*. 61:1 (2012) 162-172.
- Harken, S. E. et al. (2006). Subject semantic interoperability: final report. Report of the Subcommittee on Semantic Interoperability to the ALCTS Subject Analysis Committee. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.7413&rep=rep1&type=pdf> (2015-12-20).
- Heron, S. J., Simpson, B., Weiss, A. K., & Phillips, J. (2013). Merging catalogs: creating a shared bibliographic environment for the State University Libraries of Florida. // *Cataloging & Classification Quarterly*. 51:1-3 (2013) 139-155.
- IFLA (2010). Functional Requirements for Subject Authority Data (FRSAD): a Conceptual Model. <http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf> (2015-12-20).
- Keller, C. D. (1974). Union catalogs and lists: aspects of national and California coverage. // *Occasional papers* (University of Illinois at Urbana-Champaign. Graduate School of Library Science). 114 (1974).
- Kilgour, F. G. (1970). History of library computerization. // *Journal of Library Automation*. 3:3 (1970) 218-229.
- Kilgour, F. G., Long, P. L., Landgraf, A. L., & Wyckoff, J. A. (1972). The shared cataloging system of the Ohio College Library Center. // *Journal of Library Automation*. 5:3 (1972) 157-183.
- Kohl, D. F. (1997). Resource sharing in a changing Ohio environment. // *Library Trends*. 45:3 (1997) 435-447.
- Koltay, K. (2004). Subject cataloging in a cooperative cataloging environment: a case study. // Lass, A.; Quandt, R. E. (Eds.). *Union catalogs at the crossroad*. Hamburg: Hamburg University Press, 2004. 305-325.
- Kuncaitis, Y. (1968). Union catalogs and bibliographic centers: a state-of-the-art review. Columbus: The State Library of Ohio, 1968.
- Lancaster, F. W., Connell, T. H., Bishop, N., & McCowan, S. (1991). Identifying barriers to effective subject access in library catalogs. // *Library Resources and Technical Services*. 35:4 (1991) 377-391.
- Lancaster, F. W. (2002). *El control del vocabulario en la recuperación de información*. 2ª ed. Valencia: Universitat de Valencia, 2002.
- Landry, P. (2004). Multilingual subject access: the linking approach of MACS. // *Cataloging & Classification Quarterly*. 37:3-4 (2004) 177-191.
- Lass, A. & Quandt, R. E. (Eds.). (2004). *Union catalogs at the crossroad*. Hamburg: Hamburg University Press, 2004.
- Lazarinis, F. (2015). *Cataloguing and classification: an introduction to AACR2, RDA, DDC, LCC, LCSH and MARC 21 standards*. Waltham, MA: Chandos Publishing, 2015.
- Lemos, L. & Macedo, M. C. (2003). *A Cooperação entre as Bibliotecas do Ensino Superior em Portugal: passado, presente e perspectiva futura*. // 9ª Jornadas Porbase. Lisboa, 2003.
- Lynch, C. A. (1997). Building the infrastructure of resource sharing: union catalogs, distributed search, and cross-database linkage. // *Library trends*. 45:3 (1997) 448-461.
- Martins, M. C. B. (2014). *Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal: o sistema no Instituto Politécnico de Portalegre*. Salamanca: Ed. Universidad de Salamanca, 2014. Tese de Doutoramento.
- Mendes, M. T. P., & Simões, M. G. (2002). *Indexação por assuntos: princípios gerais e normas*. Lisboa: Gabinete de Estudos a&b, 2002.
- OCLC (2015). OCLC prints last library catalog cards. <https://www.oclc.org/news/releases/2015/201529dublin.en.html> (2015-12-20).
- Olson, H. A., & Boll, J. J. (2001). *Subject analysis in online catalogs*. 2nd ed. Englewood, Colorado: Libraries Unlimited, 2001.
- Rebiun (2014). Manual de procedimiento para la inclusión de registros en el catálogo colectivo de REBIUN y para el uso de Z-39.50. http://www.rebiun.org/documentos/Documents/GTCATALOGO/Manualdeprocedimiento_CatalogoColectivo_enero2014.pdf (2015-12-20).
- Reitz, J. M. (2013). Union catalog. // ODLIS: Online Dictionary for Library and Information Science. http://www.abc-clio.com/ODLIS/odlis_about.aspx (2015-12-20).
- Smith, I. A. (2003). Union catalogues. // Feather, J.; Surges, P. (Ed.). *International Encyclopedia of Information and Library Science*. 2nd ed. London: Routledge, 2003. 643.
- State University Libraries of Florida (2010). Guidelines and procedures for the shared bibliographic catalog. <http://csul.net/sites/csul.fcla.edu/uploads/SULGuidelinesandProceduresforSharedBibCatalog.pdf> (2015-12-20).
- Wiley, L. (2011). Access-Related Reference Services. // Bopp, R. E.; Smith, L. C. (Eds.). *Reference and information services: an introduction*. 4th ed. ABC-CLIO, 2011.
- Wright, A. (2014). *Cataloging the world: Paul Otlet and the birth of the information age*. Oxford: Oxford University Press, 2014.
- Zeng, M. L. & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. // *Journal of the American Society for information science and technology*. 55:5 (2004) 377-395.

Copyright: © 2017. Revez. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

Received: 2016-03-11 Revised: 2016-07-26 Accepted: 2016-12-19

APÊNDICES

Apêndice01: Elaboração de projetos de implementação de catálogos coletivos



Apêndice 2. Modelos detetados na literatura acerca da relação entre os catálogos coletivos e a indexação por assuntos

